CENTRE DE ROCQUENCOURT

Institut National
de Recherche
en Informatique
et en Automatique

Domaine de Voluceau
Rocquencourt
B.P.105
78153 Le Chesnay Cedex
France
Tél. : (1) 39 63 55 11

Rapports de Recherche

N° 475

AN ACTIVE SET METHOD FOR SOLVING LINEARLY CONSTRAINED NONSMOOTH OPTIMIZATION PROBLEMS

Eliane PANIER

Janvier 1986

AN ACTIVE SET METHOD FOR SOLVING LINEARLY CONSTRAINED NONSMOOTH OPTIMIZATION PROBLEMS.

Eliane PANIER

ABSTRACT.

In this paper, we present an algorithm for solving linearly constrained problems. The method is based on two ingredients : the direction is computed by a bundle principle and the constraints are treated through an active set strategy.

We point out the difficulties which arise when the objective function is nonsmooth, in particular concerning the choice of a constraint to relax.

We show that some non degeneracy hypothesis is necessary to obtain convergence.

Key words : nonsmooth optimization, bundle algorithms, linear constraints.

RESUME.

Dans cet article, nous présentons un algorithme de résolution des problèmes avec des contraintes linéaires. Deux aspects sont pris en compte pour le développement de la méthode : la direction est calculée par une méthode de faisceau, tandis que les conditions sont traitées avec une stratégie d'ensemble actif.

Nous mettons en évidence les difficultés qui apparaissent quand la fonction d'objectif n'est pas régulière, en particulier pour le choix de la condition à relaxer.

Nous montrons que la convergence est nécessaire avec une certaine hypothèse "de régularité".

Mots clés : optimization non différentiable, algorithme de faisceau, contraintes linéaires.

I. INTRODUCTION.

Given some function $f : \mathbb{R}^n \to \mathbb{R}$ one is interested in solving the following problem (P)

$$(P) \qquad \begin{cases} \text{minimize } f(x) \\ \text{subject to } Ax \leq b \\ \qquad\qquad x \in \mathbb{R}^n \end{cases}$$

where b is an m-vector and A an $m \times n$ matrix.

The so-called active set algorithms constitute a large class of methods for handling this kind of problem. A typical iteration of these methods can be described as follows. Given a feasible point x_k (i.e. a point such that $Ax_k \leq b$) :

1) identify the set $\mathcal{I}_k = \{i \mid (Ax_k)_i = b_i\}$ of active constraints ;

2) compute a descent direction d_k which satisfies the property $\mathcal{A}_k d_k = 0$ where $\mathcal{A}_k$ is the submatrix of A constituted by the rows of A corresponding to $\mathcal{I}_k$;

3) relax one or several constraints if some relaxing rule allows it. If it is the case, compute a new direction d_k as above ;

4) perform a line search along d_k with maximal step $\theta_k = \max \{t \mid A(x_k + t\, d_k) \leq b\}$ in order to get a new feasible point $x_{k+1} = x_k + t_k d_k$.

Many methods have been proposed. They differ by the choice of a descent direction and the relaxing rule. The gradient projection method of ROSEN [16] has been the first method in this area. The descent direction d_k is computed as the opposite of the projection of the gradient on the subspace of active constraints, namely

$$d_k = - P_{\mathcal{A}_k} \nabla f(x_k) \text{ where } P_{\mathcal{A}_k} = I - \mathcal{A}_k^t (\mathcal{A}_k \mathcal{A}_k^t)^{-1} \mathcal{A}_k .$$

If $| d_k |$ is small i.e. if the function f is approximately minimized in the subspace $\{x \mid \mathcal{A}_k x = b\}$ then a first order estimate M_k of the Kuhn-Tucker multipliers : $M_k = -(\mathcal{A}_k \mathcal{A}_k^t)^{-1} \mathcal{A}_k \nabla f(x_k)$ is computed. If all the components of M_k are non negative then x_k is an approximate minimum. Otherwise, the constraint corresponding to the most negative component of M_k is relaxed. In order to get a new feasible direction, only one constraint can be deleted.

In other methods, as for example in [3], [4] and [9] a descent direction is computed as the solution of a quadratic programming problem (Q) associated with f

$$(Q) \quad \begin{array}{l} \text{minimize } \nabla f(x_k)^t d + \tfrac{1}{2} d^t H_k d \\ \\ \text{subject to } \mathcal{A}_k d = 0. \end{array}$$

Here H_k denotes an approximation of the Hessian matrix of f at x_k. For the relaxing rule, they use, as approximate multipliers, the Kuhn-Tucker multipliers associated with the solution of the quadratic problem (Q).

Active set methods are subject to zigzagging or jamming phenomenon : the iterates oscillate between several subspaces (see [20]). It is the case, for example, for the method initially described by ROSEN. Several ways of preventing such phenomenon have been considered. Mc CORMICK in [10] does not restrict the step in the line search to be of the form $x_k + t_k d_k$ with $t_k \leq \theta_k$ but rather performs a line search on the function $f(P_{\mathcal{A}_k}(x))$. ε-active set methods are another way of avoiding zigzag. The set $\mathcal{I}_k$ of active constraints at x_k is replaced by the set $\mathcal{I}_k^\varepsilon = \{i \mid (Ax_k)_i \geq b - \varepsilon\}$ where ε is a positive tolerance. This approach is used, for example, in [17]. Finally, a third way of keeping out of zigzag is to refine the relaxing rule. Rather complicated relaxing rules have been established in [21], [14] and [3]. BYRD and SHULTZ, in [1], define a general relaxing rule and show that a large class of algorithms using it are convergent. Their relaxing rule is not too restrictive but allows only to eliminate one constraint at once. DEMBO and SAHI [2] have proposed a more flexible relaxing rule which authorizes to delete several constraints simultaneously.

The first two approaches can be generalized easily when the objective function is nonsmooth. For example, SREEDHARAN [18] extends the method developed in [17] to solve problem (P) with a special nonsmooth objective function. On the other hand, it is quite easy to generalize the ε-active set method to the nondifferentiable case. A particular algorithm, in nonsmooth optimization, working with ε-active set is given in [12]. An everlasting difficulty with these methods is the choice of ε : which suitable ε to choose ? How to update ε ?...

In this paper, we are interested in solving problem (P) when the objective function f is convex but not necessarily differentiable. Convexity, which is not essential for our study, is assumed only for simplicity. We present and prove the convergence of an algorithm which generalizes the gradient projection method of ROSEN with a relaxing rule inspired by the study done in [1].

The main difficulty encountered in the nonsmooth case is the following. At each point, one has at our disposal a set of approximate multiplier vectors and non longer, as for the differentiable case, a single multiplier vector. This leads to some trouble when defining the relaxing rule. First, it is sometimes necessary to relax several constraints simultaneously. The second problem is how to choose constraints to delete by examining the sign of the components of the multiplier vectors.

In section 2, we introduce the notations used in this paper.

We define, in section 3, an ℓ-stationary point and an algorithm to find such a point.

The study of the approximate multipliers is done in section 4.

Finally, section 5 is devoted to convergence theorems.

II. SOME DEFINITIONS AND NOTATIONS.

In this paper, we work on the space $\mathbb{R}^n$ equipped with its scalar product $\langle .,. \rangle$ and its associated norm $|\ .\ |$. The origin of $\mathbb{R}^n$ is denoted by $\underline{0}$.
An n-vector x is said to be non negative if all its components are non negative. Given a set S, the writing $|\ S\ | = q$ means that S contains q elements.

Likewise, given a q×n matrix B, we denote by $|B|$ the number of its rows ($|B| = q$). For i, $1 \le i \le q$ the row i of B is represented by the transposed n-vector $(B)_i$. We introduce the projection matrix P_B associated with B, $P_B = I - B^t (B B^t)^{-1} B$, provided B is of full row rank.

We say that some matrix C is contained (resp. strictly contained) in B and we note $C \subset B$ (resp. $C \subsetneq B$) if all the rows of C are rows of B (resp. and B is different from C).

We denote by X the feasible set of problem (P) : $X = \{x \mid Ax \le b\}$. Given some n-vector x and some non negative number ℓ, we define, (see [15]) the ℓ-subdifferential evaluated at x

$$\partial_\ell f(x) = \{g \in \mathbb{R}^n \mid f(y) \ge f(x) + \langle g, y-x\rangle - \ell \;\forall\, y \in \mathbb{R}^n\}$$

The subdifferential $\partial_0 f(x)$ is simply noted by $\partial f(x)$. An element of $\partial f(x)$ is called a subgradient. In what follows, we suppose that a function $g : \mathbb{R}^n \to \mathbb{R}^n$ is given whose value at any point x is an element of $\partial f(x)$. By means of the function g, we define a "weight" function $p : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^+$ by :

$$\forall x, y \in \mathbb{R}^n \quad p(x,y) = f(x) - f(y) - \langle g(y), x-y\rangle.$$

The value of p indicates the error obtained at x by linearizing f around y.

The algorithm described in section 3 builds a sequence $\{x_k\}_{k\in\mathbb{N}}$. The set of active indices $\mathcal{J}_k = \{i \mid (Ax_k)_i = b_i\}$ at the iterate x_k is divided in two subsets : the set I'_k of indices associated with relaxed constraints and the set I_k of remaining active indices. The submatrix $\mathcal{A}_k$ composed of the rows $(A)_i$ for i belonging to $\mathcal{J}_k$ is divided accordingly into $\mathcal{A}_k = \begin{bmatrix} A_k \\ A'_k \end{bmatrix}$.

III. AN ACTIVE SET METHOD.

In this section, we define ℓ-stationary points and ℓ-multiplier vectors and we describe an algorithm for finding ℓ-stationary points.

Some positive tolerance ℓ on the results is introduced for numerical reasons just as in [8].

3.1. Some definitions.

In what follows, we consider a point x^* belonging to X, the $m^* \times n$ submatrix A^* of its active constraints and a tolerance $\ell > 0$.

Let $\bar{A}$ be an $\bar{m} \times n$ submatrix of A^*. The point x^* is said to be an <u>ℓ-critical point associated with $\bar{A}$</u> if there exist vectors $M \in \mathbb{R}^{\bar{m}}$ and $g \in \partial_\ell f(x^*)$ satisfying the following equality

$$(3.1) \qquad g + \bar{A}^t M = 0.$$

The point x^* is said to be <u>ℓ-critical</u> if it is an ℓ-critical point associated with some submatrix $\bar{A}$ of A^*.

If x^* is an ℓ-critical point associated with a submatrix $\bar{A}$ of A^*, the vector M which appears in relation (3.1) is called an <u>ℓ-multiplier vector associated with $\bar{A}$ and g</u>. When the submatrix $\bar{A}$ is of full row rank, the ℓ-multiplier vector M associated with g and $\bar{A}$ is unique and given by

$$M = -(\bar{A}\,\bar{A}^t)^{-1}\,\bar{A}\,g.$$

More generally, if x^* is an ℓ-critical point associated with a full row rank submatrix $\bar{A}$, the evaluation at x^* of the <u>ℓ-multiplier vectors multi-function associated with $\bar{A}$</u>, $\mathcal{M}_{\bar{A}}$ is defined by

$$(3.2) \qquad \mathcal{M}_{\bar{A}}(x^*) = \{-(\bar{A}\,\bar{A}^t)^{-1}\,\bar{A}\,g \mid g \in \partial_\ell f(x^*) \text{ and } g \text{ satisfies } (3.1)\}.$$

The point x^* is said to be <u>ℓ-stationary</u> if there are some $\bar{m} \times n$ submatrix $\bar{A}$ of A^* and some non negative $\bar{m}$ vector M such that :

(i) $\quad x^*$ is critical for $\bar{A}$;

(ii) $\quad M \in \mathcal{M}_{\bar{A}}(x^*)$.

Note that, an ℓ-stationary point as presented above is nothing else than a special ℓ-solution of problem (P) as introduced in [19]. In particular, an ℓ-stationary point $x^* \in X$ satisfies :

$$f(x^*) \leq f(x) + \ell \quad \forall x \in X.$$

3.2. The algorithm.

As pointed out in the introduction, the principal steps performed by an active set algorithm are : (i) define a descent direction, (ii) relax one or more constraints leading to the computation of a new direction if some relaxing rule allows it and (iii) perform a line search along the direction.

The scheme of our algorithm is slightly different in that there is a possible loop between steps (i) and (ii). A sequence $\{x_k\}_{k \in \mathbb{N}}$ is built as follows. Let $x_k \in X$ be given.

1. Determine the set of active indices $\mathcal{I}_k = \{i \mid (Ax_k)_i = b_i\}$. Initialize the set of indices of relaxed constraints $I'_k = \emptyset$ and the set of indices of remaining active constraints $I_k = \mathcal{I}_k$. Identify the active submatrix $\mathcal{A}_k = \begin{bmatrix} A_k \\ A'_k \end{bmatrix}$.

2. Compute a descent direction d_k satisfying : $A_k\, d_k = 0$ and $A'_k\, d_k \leq 0$.

3. Relax some constraint i_0 if a relaxing rule allows it ; update $I_k \leftarrow I_k / \{i_0\}$, $I'_k \leftarrow I'_k \cup \{i_0\}$ and go back to step 2. Otherwise, go to step 4.

4. Check for optimality.

5. Perform a line search along the direction d_k to obtain a new feasible point.

□

Let us suppose that at an iterate x_k we have at our disposal the submatrix $\mathcal{A}_k$ of active constraints divided into $\mathcal{A}_k = \begin{bmatrix} A_k \\ A'_k \end{bmatrix}$. One way of generalizing ROSEN's approach, in the nonsmooth case, is to project the origin of $\mathbb{R}^n$ on the convex set $-P_{A_k}(\partial_\ell f(x_k))$ and to take that projection as a descent direction. Unfortunately, it is generally not possible to determine the ℓ-subdifferential $\partial_\ell f(x_k)$ explicitly and to compute the projection. The set $\partial_\ell f(x_k)$ is therefore replaced by a convex set G_k obtained from subgradients $g_i = g(y_i)$ $i = 0,\dots k$ accumulated successively by the algorithm

$$G_k = \{ \sum_{i=0}^{k} \lambda_i g_i \mid \lambda_i \geq 0 \ i = 0,\dots k \ ; \sum_{i=0}^{k} \lambda_i = 1 \ ; \sum_{i=0}^{k} \lambda_i \, p(x_k,y_i) \leq \ell \}.$$

On the other hand, as will be shown further, it is sometimes necessary, in the nonsmooth case, to relax several constraints simultaneously. In order to be sure that the direction computed is feasible relatively to the relaxed constraints, the gradient of deleted constraints are introduced in the computation of a descent direction as done by SCHULTZ in [17] for the differentiable case.

Finally, a feasible search direction d_k at x_k is obtained by solving the quadratic problem (3.3)

$$\underset{\lambda,\ \nu}{\text{minimize}} \ \tfrac{1}{2} \left| P_{A_k} \left(\sum_{i=0}^{k} \lambda_i g_i + \sum_{j \in I'_k} \nu_j (A)_j^t \right) \right|^2$$

$$\text{subject to} \quad \sum_{i=0}^{k} \lambda_i = 1$$

$$\lambda_i \geq 0 \quad i = 0,\dots k$$

$$\text{(3.3)} \qquad \nu_j \geq 0 \quad j \in I'_k$$

$$\sum_{i=0}^{k} \lambda_i \, p(x_k,y_i) \leq \ell$$

and setting $d_k = -P_{A_k}(g'_k + \sum_{j \in I'_k} (\nu_k)_j (A)_j^t)$ with $g'_k = \sum_{i=1}^{k} (\lambda_k)_i g_i$ where $(\lambda_k)_i$ $i = 0,\dots k$ and $(\nu_k)_j$ $j \in I'_k$ are solutions of (3.3).

Let us note that the quadratic problem (3.3) of computing a direction d_k can be easily solved by the method given by MIFFLIN in [11].

Once the search direction is computed, if non additional relaxing occurs, a line search is performed along that direction. We use here the line search described in [6] which extends the one of WOLFE for the nonsmooth case. The active set method generating feasible points, the step taken by the line search is limited to the maximal feasible step $t_k^1 = \max \{t | A(x_k + t\, d_k) \leq b\}$. Another bound $t_k^2 > 0$ computed from approximated multipliers is imposed which intends to limit the step in the vicinity of a critical but non stationary point, in accordance with what is done in [1]. Given the bound $\theta_k = \min \{t_k^1, t_k^2\}$, a negative number v_k approximating $f'(x_k, d_k)$ and three positive scalars m_1, m_2, m_3 satisfying $m_1 < m_2 < 1$ and $m_2 + m_3 < 1$, the line search proceeds as follows.

If the approximation G_k of the ℓ-subdifferential is good enough, the direction computed is a descent direction and, as in WOLFE's method, the line search furnishes a step $t_k \leq \theta_k$ satisfying

(3.4) $$f(x_k + t_k\, d_k) \leq f(x_k) + m_1\, t_k\, v_k$$

and either

(3.5) $$(g(x_k + t_k\, d_k),\ d_k) \geq m_2\, v_k \text{ (serious step)}$$

or

(3.6) $$t_k = \theta_k \text{ (maximal step).}$$

Otherwise, if the approximation of the subdifferential is not good, the step t_k performed by the line search satisfies (3.5) and

(3.7) $$p(x_k,\ x_k + t_k\, d_k) \leq m_3\, \ell.$$

The subgradient $g(x_k + t_k\, d_k)$ will enrich the approximation G_k of the subdifferential $\partial_\ell f(x_k)$.

It remains to define a relaxing rule which will lead to a convergent algorithm.

The relaxing rule which extends the one proposed by ROSEN is based on approximated multipliers

$$(3.8) \qquad M_k = -(A_k A_k^t)^{-1} A_k \left(\sum_{i=0}^{k} (\lambda_k)_i g_i + \sum_{j \in I'_k} (\nu_k)_j (A)_j^t\right)$$

where the numbers $(\lambda_k)_i$ $i = 0,\ldots k$ and $(\nu_k)_j$ $j \in I'_k$ are solutions of the quadratic problem (3.3). If some relaxing occurs, only a constraint of negative multiplier is deleted. The relaxing rule is steered by the one described in [1]. First a constraint must not be eliminated too often in view of avoiding zigzag. In short, it is better not to relax if some constraint has been added at the previous iteration. On the other hand, a relaxing has always to be done in a neighbourhood of a critical but non stationary point in order to force the multiplier vector components to be of the good sign (here, the neighbourhood is defined by $|d_k| \le \underline{d}$ where $\underline{d}$ is some given positive number). Moreover, only one relaxing may happen at a point in BYRD and SHULTZ's rule, In the nonsmooth case, one is sometimes forced to relax one constraint even if the BYRD and SHULTZ's rule does not allow it. We then relax if the search direction d_k becomes arbitrarily small, say if the following inequality holds ; $|d_k| \le \alpha_k$ where α_k is a positive but decreasing number tending to zero if an infinite number of eliminations occur.

To summarize, the relaxing rule is given by the following scheme : relax some constraint i_0 such that $(M_k)_{i_0} = \min \{(M_k)_i \mid i \in I_k\}$ if and only if the two following conditions are fulfilled

(i) $(M_k)_{i_0} < 0$;

(ii) either $|d_k| \le \alpha_k$

or

$|d_k| \le \underline{d}$ and
- no constraint was added at the arrival at x_k ;
- no constraint has already been deleted (i.e. $I'_k = \emptyset$).

If some relaxing occurs, diminish α_k to force it toward 0.

The complete algorithm can now be described.

Let be given some positive tolerances : ℓ, $0 < \beta < 1$; c ; $\underline{d}$; α ; m_1, m_2, m_3 with $m_1 < m_2 < 1$; $m_2 + m_3 < 1$ and a feasible starting point $x \in X$.

The algorithm proceeds as follows.

Step 0. Initialize : $x_0 \leftarrow x$, $y_0 \leftarrow x_0$, $\alpha_0 \leftarrow \alpha$, $g_0 \leftarrow g(y_0)$, $k \leftarrow 0$.

Step 1. Define the active set at x_k : $\mathcal{I}_k = \{i \mid (Ax_k)_i = b_i\}$ and the active submatrix $\mathcal{A}_k$ of A. Set : $I_k = \mathcal{I}_k$ and $I'_k = \emptyset$.

Step 2. Compute a descent direction d_k by solving problem (3.3).

Step 3. Determine the approximate multipliers vector M_k given by (3.8) and identify

$$v_k = - | d_k |^2 - S_k \, \ell$$

where S_k is the multiplier associated with the last constraint of (3.3).

We refer to [8] for an interpretation of v_k as $f'(x_k, d_k)$.

Step 4. If some constraint i_0 is to be deleted according to the relaxing rule defined above update the indices sets : $I_k \leftarrow I_k / \{i_0\}$; $I'_k \leftarrow I'_k \quad \{i_0\}$; diminish α_k : $\alpha_k \leftarrow \beta \, \alpha_k$ and go back to step 2. Otherwise, go to step 5.

Step 5. Check for optimality. If $| d_k | = 0$ stop (x_k is an ℓ-stationary point).

Step 6. Compute the maximal feasible step $t_k^1 = \max \{t \mid A(x_k + t \, d_k) \leq b\}$ and the limiting number

$$t_k^2 = \begin{cases} \dfrac{c}{\left| \min\limits_i \{(M_k)_i \mid (M_k)_i < 0\} \right|} & \text{if } (M_k)_i < 0 \text{ for some } i \\ + \infty \text{ otherwise} & \end{cases}$$

and perform a line search along the segment line $[0, \Theta_k]$ with $\Theta_k = \min \{t_k^1, t_k^2\}$ to obtain a step t_k satisfying either (3.4) and (3.5) (serious step) or (3.4) and (3.6) (maximal step) or (3.5) and (3.7) (null step). Compute $y_{k+1} = x_k + t_k \, d_k$ and $g_{k+1} = g\,(y_{k+1})$.

If a serious step or a maximal step has been performed : define $x_{k+1} = y_{k+1}$; update $k \leftarrow k+1$; go back to step 1.

If a null step has been obtained by the line search : define $x_{k+1} = x_k$, $I_{k+1} = I_k$ and $I'_{k+1} = I'_k$; update $k \leftarrow k+1$; go back to step 2.

3.3. Some assumptions and notations.

In this paper, we suppose that the following assumptions hold :

- The function f is convex ;
- the feasible set X is not void ;
- the set $\{x \mid f(x) \leq f(x_0)\} \cap X$ is compact ;
- all the submatrices $\mathcal{A}_k$ encountered during the progress of the algorithm have full row rank ;
- stop at step 5 of the algorithm never occurs.

IV. STUDY OF THE SEARCH DIRECTION AND OF APPROXIMATED ℓ-MULTIPLIER VECTORS.

4.1. Feasibility of the search direction.

The first thing to do is to show that the direction d_k is feasible with respect to the active constraints in order to avoid $t^1_k = 0$ at step 6 of the algorithm.

Proposition 4.1.

The direction d_k computed by the algorithm satisfies

(4.1) $$<(A)^t_j, d_k> \leq 0 \quad \forall\, j \in I_k \quad I'_k \text{ (with equality for } j \in I_k)$$

and

(4.2) $$<d_k, g_i> - S_k\, p(x_k, y_i) \leq V_k \quad i = 0,\dots,k.$$

Proof.

i) As the direction is computed from subgradients projected on the subspace $\{d \mid A_k\, d = 0\}$ we have $A_k\, d_k = 0$.

Moreover, optimality conditions associated with the quadratic problem (3.3) can be written as

$$\exists j, S_k \geq 0, \quad (\lambda_k)_i \geq 0 \quad i = 0,\dots,k, \quad (\nu_k)_j \geq 0 \quad j \in I'_k \text{ s.t.}$$

$$d_k = - \sum_{i=1}^{k} (\lambda_k)_i \, P_{A_k} g_i - \sum_{j \in I'_k} (\nu_k)_j \, P_{A_k} (A)_j^t.$$

(4.3) $$\langle -d_k, P_{A_k} g_i \rangle + S_k \, p(x_k, y_i) + \gamma \geq 0 \quad i = 0,\dots k$$

(with equality for i such that $(\lambda_k)_i \neq 0$)

(4.4) $$\langle -d_k, P_{A_k} (A)_j^t \rangle \geq 0 \quad j \in I'_k \, .$$

(with equality for j such that $(\nu_k)_j \neq 0$).

Complementarity conditions

$$S_k \sum_{i=0}^{k} (\lambda_k)_i \, p(x_k, y_i) = S_k \, \ell$$

$$\sum_{i=0}^{k} (\lambda_k)_i = 1.$$

Now, as proved in [8, page 267] for the projection matrix P_{A_k} it holds :

$$\langle d_k, P_{A_k} g \rangle = \langle d_k, g \rangle \quad \forall \, g \in \mathbb{R}^n.$$

and, in view of (4.4) we then obtain

(4.5) $$\langle d_k, (A)_j^t \rangle \leq 0 \; \forall \; j \in I'_k.$$

ii) Multiplying each inequality of type i in (4.3) by $(\lambda_k)_i$ and each inequality of type j in (4.4) by $(\nu_k)_j$ and adding, we obtain the expression of the number j as :

$$j = - |d_k|^2 - S_k \, \ell.$$

Replacing j by its value in (4.3) we get (4.2). □

4.2. Study of approximated ℓ-multiplier vectors.

The aim of this study is to show that, under some non degeneracy assumption, after relaxing in the vicinity of a critical but non stationary point, the direction computed in the new subspace has a norm bounded from below by a positive number.

4.2.1. Study of the continuity of the ℓ-multiplier vectors multi-fonction.

We establish two preliminary lemmas.

Lemma 4.2.

Let $d_k = - P_{A_k} (g'_k + \sum_{j \in I'_k} (\nu_k)_j \, (A)_j^t)$ be a direction computed by the algorithm with $g'_k = \sum_{i=0}^{k} (\lambda_k)_i \, g_i$.

Then, the vector g'_k belongs to $\partial_\ell f(x_k)$.

Proof.

See [19].

□

Lemma 4.3.

There exist positive numbers $\bar{d}$ and $\bar{M}$ such that for each $k \in \mathbb{N}$ there holds : $|d_k| \leq \bar{d}$ and $|M_k| \leq \bar{M}$ (where $|M_k|$ means the euclidian norm of M_k in $\mathbb{R}^{|I_k|}$).

Proof.

We suppose, without loss of generality, that the indices sets I_k and I'_k keep fixed values I and I'. The active submatrix is divided into $\mathcal{A}_k = \begin{bmatrix} A_I \\ A_{I'} \end{bmatrix}$.

(i) For $k \in \mathbb{N}$, the direction d_k obtained by solving the quadratic programming problem is given by :

$$d_k = - P_{A_I} \left(g'_k + \sum_{j \in I'} (\nu_k)_j \, (A)_j^t\right) \text{ with } g'_k = \sum_{i=0}^{k} (\lambda_k)_i \, g_i .$$

It obviously holds

(4.6) $$| d_k | \leq | P_{A_I} \, g'_k | .$$

Moreover, in view of lemma 4.2 we have

(4.7) $$g'_k \in \partial_\ell \, f(x_k).$$

As the sequence $\{x_k\}_{k \in \mathbb{N}}$ is bounded by hypothesis and the ℓ-subdifferential is locally bounded (see [7]) we deduce from (4.6) and (4.7) that the sequence $\{d_k\}_{k \in \mathbb{N}}$ is bounded.

(ii) It thus remains to show that the sequence of approximated multiplier vectors is bounded.

Suppose that there exists some index i_0 and some subsequence $\{M_k\}_{k \in N'} \subset \mathbb{N}$ such that

$$\{| (M_k)_{i_0} |\} \underset{\substack{k \to \infty \\ k \in N'}}{\to} \infty .$$

For k belonging to N', the search direction can be rewritten

(4.8) $$d_k = - g'_k - \sum_{j \in I'} (\nu_k)_j \, (A)_j^t - \sum_{i \in I} (M_k)_i \, (A)_i^t .$$

Dividing (4.8) by $(M_k)_{i_0}$ we obtain

(4.9) $$\frac{d_k}{(M_k)_{i_0}} = - \frac{g'_k}{(M_k)_{i_0}} - \sum_{j \in I'} \frac{(\nu_k)_j}{(M_k)_{i_0}} \, (A)_j^t - \sum_{i \in I/\{i_0\}} \frac{(M_k)_i}{(M_k)_{i_0}} \, (A)_i^t - (A)_{i_0}^t$$

Let us set

$$v_k = -\sum_{j\in I'} \frac{(\nu_k)_j}{(M_k)_{i_0}} (A)_j^t - \sum_{i\in I/\{i_0\}} \frac{(M_k)_i}{(M_k)_{i_0}} (A)_i^t .$$

As the sequence $\{d_k\}_{k\in N}$ and $\{g'_k\}_{k\in N}$ are bounded we have by taking the limit into (4.9)

$$\lim_{\substack{k\to\infty \\ k\in N'}} v_k = (A)_{i_0}^t$$

and since every vector v_k belongs to the subspace span $\{(A)_i^t \quad i \in I \cup I' / \{i_0\}\}$ this contradicts the fact that A_k is of full row rank. □

A necessary property to ensure convergence of algorithms in the differentiable case is the continuity of approximated multipliers at a critical point. This property is translated, in the nonsmooth case, as the upper-semi-continuity of the multi-valued approximated ℓ-multiplier vectors function at an ℓ-critical point.

Proposition 4.4.

Let us suppose that there exists some subset $N' \subset N$ and some point x^* for which the indices sets I_k and I'_k have fixed values I and I', and the following relations hold :

$$\{x_k\} \underset{\substack{k\in N' \\ k\to\infty}}{\to} x^* ;$$

$$\{d_k\} \underset{\substack{k\in N' \\ k\to\infty}}{\to} 0. \tag{4.10}$$

If we denote by $\bar{A} = \begin{bmatrix} A_I \\ A_{I'} \end{bmatrix}$ the active submatrix we can state :

(i) the vector x^* is an ℓ-critical point associated with $\bar{A}$;

(ii) there exists a subset $N'' \subset N'$ of indices, a $|I|$-vector M^* and a non negative $|I'|$-vector ν^* such that

$$\{M_k\} \underset{\substack{k \to \infty \\ k \in N''}}{\to} M^* ;$$

$$\begin{bmatrix} M_*^* \\ \nu \end{bmatrix} \in \mathcal{M}_{\bar{A}} (x^*).$$

Proof.

The direction associated with some k belonging to N' is given by

$$(4.11) \qquad d_k = - g'_k - \sum_{j \in I'} (\nu_k)_j \, (A)_j^t - \sum_{i \in I} (M_k)_i \, (A)_i^t .$$

From lemma 4.2 we have

$$(4.12) \qquad g'_k \in \partial_\ell \, f(x_k)$$

and, as the sequence $\{x_k\}_{k \in N'}$ tends to x^* the relation (4.12), the upper-semi-continuity of the ℓ-subdifferential and lemma 4.3 imply that there exist a subset $N'' \subset N'$ of indices and vectors $g^* \in \partial_\ell \, f(x^*)$ and M^* such that

$$(4.13) \qquad \{g'_k\} \underset{\substack{k \to \infty \\ k \in N''}}{\to} g^* ;$$

$$(4.14) \qquad \{M_k\} \underset{\substack{k \to \infty \\ k \in N''}}{\to} M^* .$$

As the set $\{ \sum_{j \in I'} \nu_j \, (A)_j^t , \ \nu_j \geq 0\}$ is closed, relations (4.10), (4.11), (4.13) and (4.14) exhibit the existence of scalars $\nu_j^* \geq 0$, $j \in I'$ for which the following equality hods :

$$g^* + \sum_{j \in I'} \nu_j^* \, (A)_j^t + A_I^t \, M^* = 0. \qquad \square$$

The limiting number t_k^2 used by line search is governed by the multipliers.

Proposition 4.5.

The sequence $\{t_k^2\}_{k \in \mathbb{N}}$ of limiting steps evaluated at the points $\{x_k\}_{k \in \mathbb{N}}$ is bounded from below by a positive number.

Proof.

Obvious from lemma 4.3 and the definition of the numbers $t^2{}_k$; $\square$

Proposition 4.6.

Let us consider some cluster point x^* of the sequence generated by the algorithm and some subsequence $\{x_k\}_{k \in N' \subset \mathbb{N}}$ converging to x^*.

If the following hypotheses hold :

(i) x^* is not an ℓ-stationary point ;

(ii) $\{d_k\} \underset{\substack{k \to \infty \\ k \in N'}}{\to} 0$;

then, the subsequence $\{t_k^2\}_{k \in N'}$ is bounded from above.

Proof.

We will prove this by contradiction. Let us suppose that (i) and (ii) hold and that there exists a subset $N'' \subset N'$ of indices for which the following relations are satisfied

$$I_k = I = I^1 \cup I^2, \; I_k' = I' \;;$$

$$(M_k)_i \geq 0 \quad i \in I^1, \quad (M_k)_i < 0 \quad i \in I^2 \;; \tag{4.15}$$

$$\text{either } I^2 = \emptyset \text{ or } (M_k)_i \underset{\substack{k \to \infty \\ k \in N''}}{\to} 0 \quad \forall\, i \in I^2. \tag{4.16}$$

If we denote by $\bar{A} = \begin{bmatrix} A_I \\ A_{I'} \end{bmatrix}$ the active submatrix we conclude, in view of proposition 4.4, that :

(4.17) the vector x^* is an ℓ-critical point

and there exists a subset N" N' of indices, a $|I|$-vector M^* and a non negative $|I'|$-vector ν^* such that

(4.18) $$\{M_k\} \underset{\substack{k \to \infty \\ k \in N''}}{\longrightarrow} M^* ;$$

(4.19) $$\begin{bmatrix} M_*^* \\ \nu \end{bmatrix} \in \mathcal{M}_{\bar{A}}(x^*).$$

But from (4.15), (4.16) and (4.18) we obtain

(4.20) $$(M^*)_i \geq 0 \quad i \in I^1 \quad , \quad (M^*)_i = 0 \quad i \in I^2.$$

The relations (4.17), (4.19) and (4.20) mean that the point x^* is an ℓ-stationary point which contradicts hypothesis (i).

□

<u>4.2.2. Degree of confidence to put in ℓ-multiplier vectors associated with an ℓ-critical point and non degeneracy hypothesis.</u>

In differentiable optimization, the multiplier associated with some constraint expresses the cost of the constraint (see [13]). Unfortunately the ℓ-multiplier vectors as defined in (3.2) do not have this property. On the other hand, one has sometimes to relax more than one constraint at a time, which was not the case in differentiable optimization.

To see this, consider the following example.

minimize $f(x)$

subject to $x \in \mathbb{R}^2$, $Ax \leq 0$

where A is a 2×3 matrix given by $A = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$. Let us suppose that the function f is convex and that the subdifferential associated with the point $x^* = (0,0)^t$ is the following

$$\partial f(x^*) = \{\begin{pmatrix} 0 \\ -1 \end{pmatrix}, \begin{pmatrix} -1 \\ 0 \end{pmatrix}\}.$$

Set $\ell = 0$. The vector x^* is an ℓ-stationary point associated with A and the set of ℓ-multipliers vectors at x^* is the segment :

$$\mathcal{M}_A(x^*) = [(\begin{smallmatrix} 0 \\ -1 \end{smallmatrix}), (\begin{smallmatrix} -1 \\ 0 \end{smallmatrix})].$$

If only one constraint is relaxed, the point x^* remains ℓ-stationary. But if constraints one and two are altogether eliminated the vector x^* is no longer an ℓ-stationary point. The trouble comes from the fact that the sign of multipliers associated with one constraint is not well defined. To prevent such an event we must impose some non degeneracy hypothesis.

<u>Non degeneracy.</u>

Let be $x^* \in X$ an ℓ-critical point for (P) and A^* the submatrix of its active constraints. The point x^* is said to be non degenerate if it holds

$$\bar{A} \subsetneq A^* \Rightarrow \underline{0} \notin \{P_{\bar{A}}\, g \mid g \in \partial_\ell f(x^*)\}$$

As the function f is convex, the non degeneracy hypothesis simply says that the vector x^* is an ℓ-solution (see [15]) of the problem

minimize $f(x)$

subject to $A^* x^* = b^*$

(where b^* is the restriction of b to active constraints) but that, after the elimination of any constraint, it is no more a solution of the new problem.

<u>Study of multipliers under the non degeneracy hypothesis.</u>

We show that the situation described in 4.2.2. can non more happen i.e. that the sign of components of ℓ-multipliers vectors is well defined at a non degenerate ℓ-critical point.

Proposition 4.7.

Let us suppose that x^* is a non degenerate ℓ-critical point for (P) then :

(i) the matrix A^* of active constraints at x^* is of full row rank ;

(ii) for any given active index i, the i^{th} component i of all elements of $\mathcal{M}_{A^*}(x^*)$ have the same sign denoted by sign $(\mathcal{M}_{\underline{A}^*}(x^*))_i$;

(iii) the point x^* is not an ℓ-critical point associated with a submatrix $\bar{A}$ strictly contained in A^*.

Proof.

We will show that (ii) holds by a method of contradiction. Establishing (i) can be done analogously . Property (iii) is a definition.

Let us suppose that there exist some active index i_0 and two elements of $\mathcal{M}_{A^*}(x^*)$ for which the signs of the component i_0 are not the same i.e. :

$$\exists g_1^*,\ g_2^* \in \partial_\ell f(x^*) \text{ and } M_1^*, M_2^* \in \mathbb{R}^{m^*} \text{ such that}$$

(4.21) $\qquad g_1^* + A^{*t} M_1^* = 0$

(4.22) $\qquad g_2^* + A^{*t} M_2^* = 0$

with $(M_1^*)_{i_0} < 0$ and $(M_2^*)_{i_0} > 0$.

But, in that case, there exists some scalar μ belonging to $]0,1[$ for which there holds

$$(\mu M_1^* + (1-\mu) M_2^*)_{i_0} = 0.$$

Multipling (4.21) by μ and (4.22) by $(1-\mu)$ and adding we obtain :

$$g^* + A^{*t} M^* = 0$$

where the vectors g^* and M^* are defined as

$$g^* = \mu\, g_1^* + (1-\mu)\, g_2^* \in \partial_\ell f(x^*)$$

$$M^* = \mu\, M_1^* + (1-\mu)\, M_2^*$$

and the identity $(M^*)_{i_0} = 0$ contradicts the non degeneracy hypothesis. □

The sign of components of approximated ℓ-multiplier vectors is well defined in the vicinity of a non degenerate ℓ-critical point.

Proposition 4.8.

Let us consider some cluster point x^* of the sequence generated by the algorithm and some subsequence $\{x_k\}_{k \in \mathbb{N}' \subset \mathbb{N}}$ converging to x^*. Let us suppose that the subsequence $\{d_k\}_{k \in \mathbb{N}'}$ tends to zero (which implies that x^* is an ℓ-critical point in view of proposition 4.4.). If moreover x^* is non degenerate then, for k large enough in $\mathbb{N}'$ we have :

(i) $\mathcal{A}_k = A^*$;

(ii) $\mathcal{D}_k \quad \{i \mid (\mathcal{M}_{A^*}(x^*))_i < 0\}$;

(iii) $(M_k) < 0 \quad \forall\, i$ s.t. $(\mathcal{M}_{A^*}(x^*))_i < 0$;

(iv) $(M_k)_i > 0 \quad \forall\, i \in \mathcal{D}_k$ s.t. $(\mathcal{M}_{A^*}(x^*))_i > 0$.

Proof.

The above proposition can easily be proved using propositions 4.4. and 4.7.

□

Consequence of multipliers analysis : two fundamental propositions.

The two following propositions, which appear as a consequence of multipliers analysis, are essential for convergence proofs. They ensure that, under some conditions, the search direction is not adherent to zero.

Proposition 4.9.

Let us consider some cluster point x^* of the sequence generated by the algorithm, its active submatrix A^* and some subsequence $\{x_k\}_{k \in N' \subset \mathbb{N}}$ converging to x^*. Then, there exists some positive scalar $\xi = \xi(x^*)$ such that :

(i) if x^* is not an ℓ-critical point, then for all k belonging to N' large enough we have $|d_k| > \xi$;

(ii) if x^* is a non degenerate ℓ-critical point then we have $|d_k| > \xi$ for all k belonging N' large enough for which one of the two following situations occurs.

a) $\mathcal{A}_k \subsetneq A^*$,

b) $\mathcal{A}_k = A^*$ and $\exists i$ s.t. $i \notin I_k$ and Sign $(\mathcal{M}_{A^*}(x^*))_i = -1$.

Proof.

Let us suppose that we can extract a subsequence $\{d_k\}_{k \in N'' \subset N'}$ converging to zero for which the indices sets I_k and I'_k keep fixed values I and I'.

If we denote by $\bar{A} = \begin{bmatrix} A_I \\ A_{I'} \end{bmatrix}$ the active submatrix then, in view of proposition 4.4. the point x^* is an ℓ-critical point associated with $\bar{A}$. If moreover, x^* is non degenerate then we conclude from proposition 4.8 that $\bar{A} = A^*$ and that

$$\forall i \; \{\text{Sign } (\mathcal{M}_{A^*}(x^*))_i = -1\} \Rightarrow i \in I$$

which completes the proof.

□

Proposition 4.10.

Let us consider some cluster point x^* of the sequence generated by the algorithm and a subsequence $\{x_k\}_{k \in N' \subset \mathbb{N}}$ converging to x^*. If the following hypotheses hold :

(i) $I'_k = \emptyset \quad \forall k \in N'$;

(ii) $\{d_k\} \underset{\substack{k \to \infty \\ k \in N'}}{\to} 0$;

(iii) x^* is non degenerate ;

(iv) x^* is not ℓ-stationary ;

then, for k belonging to N' large enough, if some constraint is relaxed at x_k, the new computed direction d_k^{new} satisfies $| d_k^{new} | > \xi$ (where the number ξ is defined in proposition 4.9). Moreover, only one elimination can occur.

Proof.

The proof of the first statement follows from propositions 4.8, 4.9 and the choice of the eliminated constraint.

If an infinite number of eliminations occur during the progress of the algorithm then, for k large enough we have $\alpha_k < \xi$ and the direction d_k^{new} computed after deleting satisfies $| d_k^{new} | > \xi > \alpha_k$ and, in accordance to the relaxing rule, no other elimination can occur. □

V. CONVERGENCE STUDY.

In what follows, we consider a cluster point x^* of the sequence generated by the algorithm described in 3.2. We denote by A^* its active submatrix.

5.1. Infinite number of null steps performed at a point.

We first examine the situation in which $\{x_k\}$ stops at x^* for some k and only null steps are performed from then on.

Theorem 5.1.

If there exists some iterate $x_j = x^*$ in which an infinite number of null steps is performed then x^* is an ℓ-stationary point.

Proof.

As no constraint is added in case of null step, the active set becomes a fixed set say I. Moreover, using relations (3.7) and (4.2) we can show, as done in [8], that there exists some subsequence $\{d_k\}_{k\in N'\subset \mathbb{N}}$ converging to zero. Then (see proposition 4.4.), the point x^* is an ℓ-critical point and, if no relaxing occurs, the components of approximated multipliers are of the good sign. Again, by means of proposition 4.4. we show that x^* is an ℓ-stationary point.

□

We now suppose that the sequence generated by the algorithm does not stop at some iterate, i.e. that only a finite number of null steps occur at each point.

During the convergence proofs, we will work on the subsequence of iterates in which either a serious step or a maximal step is performed, i.e. for which $x_{k+1} \neq x_k$. For simplicity and without impeding the proofs, we still index this subsequence by $\mathbb{N}$, and we will consider a subsequence $\{x_k\}_{k\in \mathbb{N}\subset \mathbb{N}}$ converging to the accumulation point x^*.

We denote by d_k^a, I_k^a, $I_k'^a$, A_k^a and $A_k'^a$ (resp. d_k^ℓ, I_k^ℓ, $I_k'^\ell$, A_k^ℓ and $A_k'^\ell$ the direction, indices sets and active submatrices computed at the arrival at a point x_k (resp. before leaving x_k). If no confusion is possible, we simply write d_k, I_k, I'_k, A_k and A'_k.

Remark that the submatrix A_k^a is simply $\mathcal{A}_k$. Futhermore, we always have $A_k^d \subset A_k^a$.

5.2. Two fundamental properties.

The convergence proof rests on two fundamental properties which are the ones used by BYRD and SHULTZ in [1]. The first property results from the fact that the step performed by the line search is bounded by the number t_k^2.

Proposition 5.2.

Let us suppose that the point x^* is not an ℓ-stationary point. Then the sequence $\{x_{k+1}\}_{k \in N'}$ converges to x^*.

Proof.

As the subset of indices N' is arbitrary it suffices to show that a subsequence of $\{x_{k+1}\}_{k \in N'}$ converge to x^*. Two cases are to be considered.

i) The sequence $\{| d_k^\ell |\}_{k \in N'}$ is bounded from below by a positive number γ. From the line search rule (3.4) we have

$$f(x_{k+1}) \le f(x_k) + m_1 \, t_k \, v_k$$

or, in view of the definition of v_k,

$$f(x_{k+1}) \le f(x_k) - m_1 \, t_k \, | d_k^\ell |^2$$

which gives

$$f(x_{k+1}) \le f(x_k) - m_1 \, \gamma \, | x_{k+1} - x_k |$$

and because the function f is bounded from below we then obtain $\{ | x_{k+1} - x_k | \} \underset{\substack{k \to \infty \\ k \in N'}}{\to} 0.$

ii) We can extract some subsequence $\{| d_k^\ell |\}_{k \in N'' \subset N'}$ converging to zero. For k belonging to N" we have

$$| x_{k+1} - x_k | = t_k \, | d_k^\ell | \le t_k^2 \, | d_k^\ell |$$

but the value of t_k^2 is bounded (see proposition 4.6) and then $| x_{k+1} - x_k | \underset{\substack{k \to \infty \\ k \in N''}}{\to} 0.$

□

The second proposition is a consequence of the convergence of the algorithm for the unconstrained problem.

Proposition 5.3.

If there exists a subsequence $\{x_k\}_{k \in N'' \subset N'}$ for which the step t_k obtained by the line search satisfies $t_k < t_k^1$ then the sequence of directions computed before leaving points of that subsequence satisfies $\{d_k^\ell\} \underset{\substack{k \to \infty \\ k \in N''}}{\to} 0$.

Proof.

The proof which uses proposition 4.5 is quite similar to the one of the theorem 2.5.3. of [8] and is not given here.

□

During convergence proofs, it will be sometimes more adapted to use proposition 5.3. expressed the following way.

Proposition 5.4.

If there exists some subset $N'' \subset N'$ of indices and some positive scalar d for which there holds

$$|d_k^\ell| \geq d \quad \forall\, k \in N''$$

then, for all k belonging to N" large enough, some constraint has to be added by the step from x_k to x_{k+1} (i.e. $t_k = t_k^1$).

5.3. One technical lemma.

The following lemma is necessary to show convergence. Its proof, which is of little use, is postponed as an appendix.

Lemma 5.5.

Let us suppose that x^* is not an ℓ-stationary point and that there exists some subsequence $\{x_k\}_{k \in N'' \subset N'}$ for which there holds :

(i) $\{d_k^{\ell}\}_{\substack{k \to \infty \\ k \in N''}} \to 0$;

(ii) $t_k < t_k^{1}$;

(iii) $\langle (A)_j^t, d_k^{\ell} \rangle = 0 \quad \forall j \in I_k'^{\ell}$;

then the subsequence $\{d_{k+1}^{a}\}_{k \in N''}$ converges to zero on N".

Remark that relation (iii) is satisfied in particular when $I_k'^{\ell} = \emptyset$.

5.4. Convergence results.

We first study the situation in which a finite number of eliminations occur.

Theorem 5.6.

Let us suppose that a finite number of eliminations occur during the progress of the algorithm then the vector x^* is an ℓ-stationary point.

Proof.

Under the stated assumptions, there exist some index K, some submatrix $\bar{A}$ of A^* and some positive scalar α for which there holds

$$\forall k \geq K \quad \alpha_k = \alpha, \ A_k = \bar{A}, \ I_k'^{\ell} = \emptyset \text{ and } t_k < t_k^{1}.$$

i) From proposition 5.3, the subsequence of directions $\{d_k^{\ell}\}_{k \in N'}$ converges to zero and then, the point x^* is an ℓ-critical point (see proposition 4.4).

ii) Let us study the signs of the multipliers computed at x^*. For k belonging to N' large enough, we have

$$|d_k| \leq \alpha_k \text{ and } I_k'^{\ell} = \emptyset$$

but according to the relaxing rule, this means that the vector M_k of approximated multipliers associated with the directions d_k^ℓ are non negative. In view of proposition 4.4, we then deduce that x^* is an ℓ-stationary point.

□

Let us now study the general case in which an infinite number of eliminations occur.

In convergence proofs, we will suppose that k is large enough in order to satisfy :

(5.1) $\alpha_k < \xi$

Two additional lemmas are needed.

Lemma 5.7.

The vector x^* is an ℓ-critical point for (P).

Proof.

We will prove this property by contradiction. Let us therefore suppose that x^* is not an ℓ-critical point for (P) and let us consider some iterate x_k, for k belonging to N' large enough. As x^* is not an ℓ-critical point for (P), using proposition 4.9 we obtain

(5.2) $| d_{k+1} | > \xi$

and from proposition 5.4, we have

(5.3) $A^a_{k+1} \supsetneq A^\ell_k$

which implies (using (5.1), (5.2), (5.3) and the relaxing rule) $A^a_{k+1} = A^\ell_{k+1}$. Therefore $A^\ell_{k+1} \supsetneq A^\ell_k$. A recurrence process, justified by proposition 5.2 leads to $|\mathcal{A}_{k+m+1}| > m$ which is impossible. Then, x^* is an ℓ-critical point for (P).

□

Lemma 5.8.

Let us suppose that x^* is not ℓ-stationary for (P). Then one of the two following situations occurs.

(i) For every subsequence $\{x_k\}_{k\in\tilde{N}} \subset N$ converging to x^* there exists some integer K such that

$$\forall\ k \in \tilde{N}\ \text{ s.t. } k \geq K, \quad \mathcal{A}_k = A.^*$$

(ii) There exists some subsequence $\{x_k\}_{k\in\tilde{N}} \subset N$

$$\forall\ k \in \tilde{N} \quad \mathcal{A}_k \underset{\neq}{\subset} A^*, \quad \mathcal{A}_{k+1} = A.^*$$

Proof.

The proof is quite analogous to that of lemma 5.7.

□

We can now establish the convergence theorem.

Theorem 5.9.

If the vector x^* is non degenerated it is an ℓ-stationary point for (P).

Proof.

From lemma 5.7. we have that x^* is an ℓ-critical point for (P). We use again a contradiction method to show that x^* is an ℓ-stationary point for (P). Let us suppose that x^* is non degenerate and that there exists some active index i for which sign $(\mathcal{M}_{A^*}(x^*))_i = -1$.

a) Let us show that we can extract a subsequence $\{x_k\}_{k\in N''\subset N}$ for which there hold :

(5.4) $$\{x_k\} \underset{\substack{k \to \infty \\ k \in N''}}{\longrightarrow} x^*;$$

(5.5) $\mathcal{A}_k = A^*$;

and either

(5.6) (a) $\forall\, k \in N'' \; |d_k^a| \leq \alpha_k$

or

(5.6) (b) $\forall\, k \in N''\; t_{k-1} < t_{k-1}^1$ and $\{d_k^a\} \underset{\substack{k \to \infty \\ k \in N''}}{\to} 0.$

Two cases are to be considered depending on the issue of lemme 5.8.

i) If situation (i) of lemma 5.8 occurs, for k belonging to N' large enough we have $\mathcal{A}_k = A^*$ and $\mathcal{A}_{k+1} = A^*$. Therefore, $t_k < t_k^1$ and, according to proposition 5.3, the subsequence $\{d_k^\ell\}_{k \in N'}$ tends to zero. On the other hand, in the present situation, any constraint eliminated before leaving x_k remains active. We then have, from lemma 5.5, $\{d_{k+1}^a\}_{k \in N'} \to 0$. And, in view of proposition 5.2, the sequence $\{x_{k+1}\}_{k \in N'}$ satisfies the appropriate properties.

ii) If situation (ii) of lemma 5.8. occurs, there exists some subsequence $\{x_k\}_{k \in \tilde{N}}$ converging to x^* for which the associated active submatrices satisfy

$$\forall\, k \in \tilde{N},\; \mathcal{A}_k \subsetneq A^*,\; \mathcal{A}_{k+1} = A^*$$

but this means that $t_k = t_k^1$ and then, we have for all k belonging to $\tilde{N}$ large enough either $|d_{k+1}^a| \leq \alpha_{k+1}$ (in which case $\{x_{k+1}\}_{k \in \tilde{N}}$ is the subsequence we are looking for) or $I_{k+1}^{\prime\,\ell} = \emptyset$. Indeed, some adding being made by the step from x_k to x_{k+1}, if $|d_{k+1}^a| > \alpha_{k+1}$ the relaxing rule does not allow to eliminate before leaving x_{k+1}. But, in the second case, we have $\mathcal{A}_{k+2} = A^*$ and $t_{k+1} < t_{k+1}^1$ and from proposition 5.3 $\{d_{k+1}^\ell\}_{k \in \tilde{N}} \to 0$ and, according to lemma 5.5, $\{d_{k+2}^a\}_{k \in \tilde{N}} \to 0$. Therefore, the sequence $\{x_{k+2}\}_{k \in \tilde{N}}$ satisfies the appropriate properties.

b) Consider the subsequence constructed in a. The subsequence of directions $\{d_k^a\}_{k\in N''}$ tends to zero (see(5.6)) and from proposition 4.8 we have, for k belonging to N" large enough

$$(5.7) \qquad (M_k)_i < 0 \text{ (resp.}>0) \ \forall i \text{ s.t. } (\mathcal{M}_{A^*}(x^*))_i < 0 \text{ (resp.}>0).$$

The relaxing rule, relations (5.6), (5.7) and the existence of a negative ℓ-multiplier show that some constraint i_0 of negative ℓ-multiplier has to be deleted before leaving the point x_k. Only one constraint may be relaxed (see proposition 4.10). Therefore,

$$I_k^{\prime\ell} = \{i_0\} \text{ where Sign } (\mathcal{M}_{A^*}(x^*))_{i_0} = -1.$$

And, according to relation (5.5), propositions 4.1, 4.10 and 5.4, there holds

$$(5.8) \qquad \mathcal{A}_{k+1} = A^* / (A)_{i_0} \cup (A)_i$$

where $(A)_i$ is some row of A which is not contained in A^*. Now, for k large enough, we have $\mathcal{A}_{k+1} \subset A^*$ which contradicts relation (5.8).

□

Acknowledgments.

The author wishes to thank C. LEMARECHAL for his many fruitful suggestions and J.J. STRODIOT for his help.

Appendix.

Proof of lemma 5.5.

Let us denote by N_k (resp. N_{k+1}) the index of the last subgradient added before computing the direction d_k^ℓ (resp. d_{k+1}^a).

i) The direction d_k^ℓ is defined as

$$d_k^\ell = - P_{A_k^\ell} \left(\sum_{i=0}^{N_k} (\lambda_k)_i \, g_i + \sum_{j\in I_k^{\prime\ell}} (\nu_k)_j \, (A)_j^t\right)$$

where $(\lambda_k)_i$ $i = 0,\dots N_k$ and $(\nu_k)_j$ $j \in I_k'^{\ell}$ are solutions of (3.3).

From hypothesis (iii) we deduce,

$$(A.1) \qquad d_k^{\ell} = - P_{\mathscr{A}_k} \left(\sum_{i=0}^{N_k} (\lambda_k)_i \, g_i + \sum_{j \in I_k'^{\ell}} (\nu_k)_j \, (A)_j^t \right)$$

ii) As $\mathscr{A}_k = \mathscr{A}_{k+1}$, we have from (4.2),

$$(A.2) \quad |d_{k+1}^a|^2 \leq S_{k+1}\, p(x_{k+1}, y_i) - \langle d_{k+1}^a, P_{\mathscr{A}_k} g_i\rangle - S_{k+1}\, \ell \quad i = 0,\dots N_{k+1}.$$

Multiplying each inequality i of (A.2) by $(\lambda_k)_i$ for $i = 0,\dots N_k$ and adding we obtain (see (A.1))

$$|d_{k+1}^a|^2 \leq S_{k+1} \sum_{i=0}^{N_k} (\lambda_k)_i \, p(x_{k+1}, y_i) - S_{k+1}\, \ell + \langle d_{k+1}^a, d_k^{\ell}\rangle$$

or

$$|d_{k+1}^a|^2 \leq S_{k+1} \sum_{i=0}^{N_k} (\lambda_k)_i \, \{p(x_{k+1}, y_i) - p(x_k, y_i)\}$$
$$+ \langle d_{k+1}^a, d_k^{\ell}\rangle + S_{k+1} \left\{ \sum_{i=0}^{N_k} (\lambda_k)_i \, p(x_k, y_i) - \ell \right\}$$

and then, in view of the last constraint of (3.3)

$$|d_{k+1}^a|^2 \leq S_{k+1} \sum_{i=0}^{N_k} (\lambda_k)_i \, \{p(x_{k+1}, y_i) - p(x_k, y_i)\}$$
$$(A.3) \qquad + \langle d_{k+1}^a, d_k^{\ell}\rangle.$$

iii) Let us show that for all positive ε, there exists some integer K for which we have,

$$(A.4) \qquad \forall\, k \in N'', \; k \geq K \; \sum_{i=0}^{N_k} (\lambda_k)_i \, \{p(x_{k+1}, y_i) - p(x_k, y_i)\} \leq \varepsilon.$$

The points y_i obtained by null steps from points x_{ℓ_i} satisfy in view of (3.7),

$$f(x_{\ell_i}) - f(y_i) - \langle g(y_i), x_{\ell_i} - y_i\rangle \leq m_3\, \ell.$$

As the vectors $g(y_i)$ belong to the sets $\partial f(y_i)$, this inequality leads to

$$g(y_i) \in \partial_\ell f(x_{\ell_i}).$$

But the points generated by the algorithm belong to a compact by assumption and, as the subdifferential is locally bounded, the vectors $g(y_i)$ are bounded. Therefore, for k belonging to N" large enough we have, from proposition 5.2 and the continuity of f

$$f(x_{k+1}) - f(x_k) - \langle g(y_i), x_{k+1} - x_k\rangle \leq \varepsilon \;\forall i = 0,\ldots N_k$$

and, by definition of the weight function, this implies

$$p(x_{k+1}, y_i) \leq p(x_k, y_i) + \varepsilon \;\forall\; i = 0,\ldots N_k.$$

Then,

$$\sum_{i=0}^{N_k} (\lambda_k)_i \{p(x_{k+1}, y_i) - p(x_k, y_i)\} \leq \varepsilon.$$

iv) From relations (A.3) and (A.4) we deduce that for all positive ε there exists an integer K such that

$$\forall\; k \in N'', k \geq K \;\; |d^a_{k+1}|^2 \leq S_{k+1}\; \varepsilon + |d^a_{k+1}|\,|d^\ell_k|.$$

But, from optimality conditions (4.2), we see that the number S_{k+1} is bounded by some positive scalar, say $\bar{S}$, and then, for all positive ε, there exists an integer K such that :

$$\forall\; k \in N'', k \geq K \;\; |d^a_{k+1}|\,(|d^a_{k+1}| - |d^\ell_k|) \leq \bar{S}\; \varepsilon$$

and this relation can only occur if the subsequence $\{d^a_{k+1}\}_{k\in N''}$ tends to zero.

□

BIBLIOGRAPHY.

[1] R.H. BYRD and G.A. SHULTZ, **"A pratical class of globally convergent active set strategies for linearly constrained optimization"**, Mathematical Programming (to appear).

[2] R.S. DEMBO and S. SAHI, **"A convergent framework for constrained nonlinear optimization"**, Yale University, Department of Mathematics, (September 1983).

[3] R. FLETCHER, **"Pratical methods of optimization"**, Vol 2., John Wiley and Sons, New-York (1981) 113-117.

[4] P.E. GILL, W. Murray and M.H. WRIGHT, **"Pratical optimization"**, Academic Press, (1981).

[5] J.B. HIRIART-URRUTY, **"ε-subdifferential calculus"**, in : Convex analysis and optimization, Research Notes in Mathematics Series, 57, Pitman Publishers, (1982).

[6] C. LEMARECHAL, **"A view of line searches"**, in : Optimization and optimal control, Editors Ausldender, Oettli and J. Stoer, Lectures Notes in Control and Information Sciences, 30, Springer-Verlag, (1981), 59-78.

[7] C. LEMARECHAL, **"Basic theory in nondifferentiable optimization. A tutorial and algorithmic oriented approach"**, INRIA, Rocquencourt, Rapport de Recherche n°181, (Décembre 1982).

[8] C. LEMARECHAL, J.J. STRODIOT and A. BIHAIN, **" On a bundle algorithm for nonsmooth optimization"**, in Nonlinear Programming, 4, Academic Press, (1981), 245-282.

[9] M. LENARD, **"A computational study of active set strategies in nonlinear programming with linear constraints"**, Mathematics Programming, 16, (1979), 81-97.

[10] G.P. Mc. CORMICK, **"Anti-zigzagging by bending"**, Mgnt Sci., 15, (1969), 315-320.

[11] R. MIFFLIN, " **A stable method for solving certain constrained least squares problems**", Mathematical Programming, 16, (1979), 141-158.

[12] V.H. NGUYEN and J.J. STRODIOT, "**A linearly constrained algorithm not requiring derivative continuity**", Eng. Struct., 6, (1984), 7-11.

[13] M.J.D. POWELL, "**Gradient conditions and Lagrange multipliers in nonlinear programming**", in : Nonlinear optimization theory and algorithms, edited by L.C.W. Dixon, E. Spedicato, G.P. Szegö, Birkaüser, (1980), 201-220.

[14] K. RITTER, "**A method of conjugate directions for linearly constrained nonlinear programming problems**", Siam. J.Number. Anal., 12, (1975), 273-303.

[15] R.T. ROCKAFELLAR, "**La théorie des sous gradients et ses applications à l'optimisation**", Collection de la Chaire Aisenstadt, Les presses de l'Université de Montréal, Canada.

[16] J.B. ROSEN, "**The gradient projection method for nonlinear programming. Part I. Linear Constraints,**" Siam Journal, 8, (1960), 181-217.

[17] H.K. SCHULTZ, "**A Kuhn-Tucker Algorithm**", Siam J. Control, 11, (1973), 438-445.

[18] V.P. SREEDHARAN, " **A subgradient projection algorithm**", Journal of Approximation Theory, 35, (1982), 111-126.

[19] J.J. STRODIOT, V.H. NGUYEN and N. HEUKEMES, "**ε-optimal solutions in nondifferentiable convex programming and some related questions**", Mathematical Programming, 25, North-Holland publishing Company, (1983), 1-22.

[20] P. WOLFE, "**On the convergence of gradient methods under constraint**", I.B.M. J. Res and Dev., 16, (1972) 407-411.

[21] ZOUTENDIJK, "**Nonlinear Programming, Computational methods**", in : J. Abadie ed., Integer and Nonlinear Programming (North-Holland Publishing Co., Amsterdam, (1970).